Uncertainties in LCA (Subject Editor: Andreas Ciroth)

Discussion Article #

New Stochastic Simulation Capability Applied to the GREET Model*

Karthik Subramanyan[1], Ye Wu[2], Urmila M. Diwekar[1] and Michael Q. Wang[2]**

[1] Vishwamitra Research Institute, Center for Uncertain Systems, Tools for Optimization and Management, Westmont, IL 60559, USA
[2] Center for Transportation Research, Argonne National Laboratory, Argonne, IL 60439, USA

** **Corresponding author** (urmila@vri-custom.org)

DOI: http://dx.doi.org/10.1065/lca2007.07.354

Please cite this paper as: Subramanyan K, Wu Y, Diwekar UM, Wang MQ (2008): New Stochastic Simulation Capability Applied to the GREET Model. Int J LCA 13 (3) 278–285

Abstract

Background, Aims and Scope. In 1995, the Center for Transportation Research (CTR) of Argonne National Laboratory (ANL) began to develop a model, called GREET (*G*reenhouse gases, *R*egulated *E*missions, and *E*nergy use in *T*ransportation), for estimating the full fuel-cycle energy and emissions impacts of alternative transportation fuels and advanced vehicle technologies. The parametric assumptions used in the GREET model involve uncertainties. A new stochastic simulation tool, developed by Vishwamitra Research Institute (VRI), is built into the GREET model to address uncertainties. This paper presents the methodology and features of this new stochastic simulation tool and evaluates the performance of the sampling techniques in the tool.

Methods. The new tool is interfaced through the graphical user interface (GUI) to perform the stochastic simulation. In general, five steps need to be followed to run a complete simulation: 1) Specify probability distribution functions; 2) Indicate the number of samples and the sampling technique; 3) Define the forecast variables; 4) Delete distribution functions (if necessary); and 5) Propagate the uncertainties and statistically analyze the outputs. The GREET model contains more than 700 default distribution functions for a wide variety of key parameters and as many as 3000 forecast variables. The stochastic simulation tool has been developed to incorporate 11 probability distribution function types for representing uncertain parameters and four sampling techniques (Monte Carlo sampling [MCS], Hammersley Sequence sampling [HSS], Latin Hypercube sampling [LHS] and Latin Hypercube Hammersley sampling [LHHS]) for stochastic simulation. To evaluate the performance of the four sampling techniques, 16 independent stochastic simulation runs were conducted in GREET and the output results were analyzed and compared.

Results and Discussion. With the same number of samples, the output distribution curve simulated by HSS is the smoothest corresponding to the highest level of uniformity. To achieve the same level of smoothness as HSS with 1,000 samples, LHHS needs to be simulated with ~1500 samples and LHS and MCS with ~3,000 samples. As a result, HSS can achieve more than 200% reduction in running time compared to LHS or MCS without compromising the accuracy and quality of the prediction curves. The simulated mean values are close enough to the actual mean value (within ±1%) despite the selection of sampling technique and the number of samples (between 1,000 and 4,000). The standard deviation values from each other are close enough as well (within ±5%). It shows the trend that the increasing number of samples makes the simulated mean value marginally closer to the actual mean value; however, the improvement effect is negligible. The simulation time is strictly positive-correlated with the number of samples; therefore, the trade-off between extending simulation time and improving the smoothness of the output distribution curve needs to be carefully assessed.

Conclusion. A new stochastic simulation tool has been developed to be built into Argonne's GREET model to enhance its capability for addressing uncertainty. This new tool guides the user in each step of the process through the user-friendly GUI windows. According to the performance comparison among the four sampling techniques, HSS was found to be the most efficient technique. Therefore, HSS was set as the default technique in GREET.

Keywords: Distribution function; GREET model; sampling technique; stochastic simulation; uncertainty; well-to-wheels analysis

Introduction

Uncertainties are inherent in life and we have learnt to deal with them by evolving cognitive heuristics and developed strategies. However, as the system becomes complicated, involving more decisions and much higher stakes, heuristics becomes obsolete and mathematical models are required (Subramanyan and Diwekar 2005a). In situations involving uncertainties, a deterministic approach to solve the problem would produce results which might not be a true reflection of reality and in such cases stochastic simulations incorporating uncertainties need to be performed (Diwekar 2003a).

The LCA community is invited for discussion contributions to this paper.
* **ESS-Submission Editor:** Dr. Andreas Ciroth (ciroth@greendeltatc.com)

In 1995, with funding from the U.S. Department of Energy (DOE), the Center for Transportation Research (CTR) of Argonne National Laboratory (ANL) began to develop a model for estimating the full fuel-cycle energy and emissions impacts of alternative transportation fuels and advanced vehicle technologies (Wang 1996). The intent was to provide an analytical tool to allow researchers to readily analyze various parametric assumptions that affect fuel-cycle energy use and emissions associated with various fuels and vehicle technologies. The model, called GREET (Greenhouse gases, Regulated Emissions, and Energy use in Transportation), calculates fuel-cycle [often called well-to-wheels, WTW] energy use in Btu/mi and emissions in g/mi for various transportation fuels and vehicle technologies. For energy use, GREET includes total energy use (all energy sources), fossil energy use (petroleum, natural gas [NG], and coal), and petroleum use. For emissions, the model includes three major greenhouse gases (GHGs) (carbon dioxide [CO_2], methane [CH_4], and nitrous oxide [N_2O]), and five criteria pollutants (volatile organic compound [VOC], carbon monoxide [CO], nitrogen oxides [NO_X], particulate matter with a diameter of 10 micrometers or less [PM_{10}], and sulfur oxides [SO_x]). Since the release of the first version of GREET (GREET1.0) in 1996, Argonne continues to update and upgrade the model. In November 2005, the most recent version – GREET1.7 – was released. The new version reflects many new efforts conducted by Argonne during the last several years, including many new features, new fuel/vehicle pathways, and up-to-date information regarding energy use and emissions for fuel production activities and vehicle operations (Wang et al. 2005). **Fig. 1** shows the GREET WTW modeling boundary. The GREET model is in the public domain, and any party can use it free of charge. The model and its associated documents are posted at Argonne's GREET website: http://www.transportation.anl.gov/software/GREET/index.html.

The GREET model incorporates a large number of input parameters and a wide variety of output results. Many of the input parameter assumptions involve uncertainties, which require probability distributions to represent the trend of occurrence of the parameter over a specific range that define the uncertainty (General Motors Corporation et al. 2001, Wang 2002, Brinkman et al. 2005, Wu et al. 2006). Since the parameters in GREET are uncertain, the resulting output variables consequently have to be represented by distributions. To address these uncertainties, a new stochastic simulation tool, developed by Vishwamitra Research Institute (VRI), is built into the GREET model. This tool has been built as a Microsoft® Excel add-in file with Visual Basic macros which can be loaded whenever the user needs to perform a stochastic simulation within the model. This new tool automates the process of setting up a stochastic simulation to a great extent and guides the user in each step of the process through the user-friendly graphical user interface (GUI) windows. It incorporates four sampling techniques including the new and efficient Hammersley Sequence Sampling (HSS) and Leaped HSS (variant of HSS) (Diwekar 2003) and an inbuilt bank of as many as 11 probability distribution function types for representing uncertain parameters. The following sections of this paper present the methodology and features of this new stochastic simulation tool, and compare the performance of the four sampling techniques for selected GREET output variables.

1 Methodology and Feature of the Stochastic Simulation Tool

This tool is interfaced through a command bar with five buttons as shown in **Fig. 2** which perform a complete stochastic simulation with the following functions:

a) **Cell Input:** Specify probability distribution functions to the input variables;
b) **Sampling:** Indicate the number of samples required and the sampling technique to be used;
c) **Forecast Cells:** Define the forecast variables;
d) **Delete Distribution:** To delete a distribution defined previously;
e) **Run Simulation:** Propagate the uncertainties and statistically analyze the outputs.

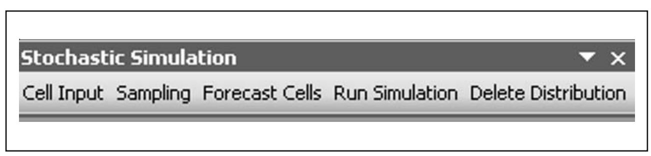

Fig. 2: Stochastic simulation command bar

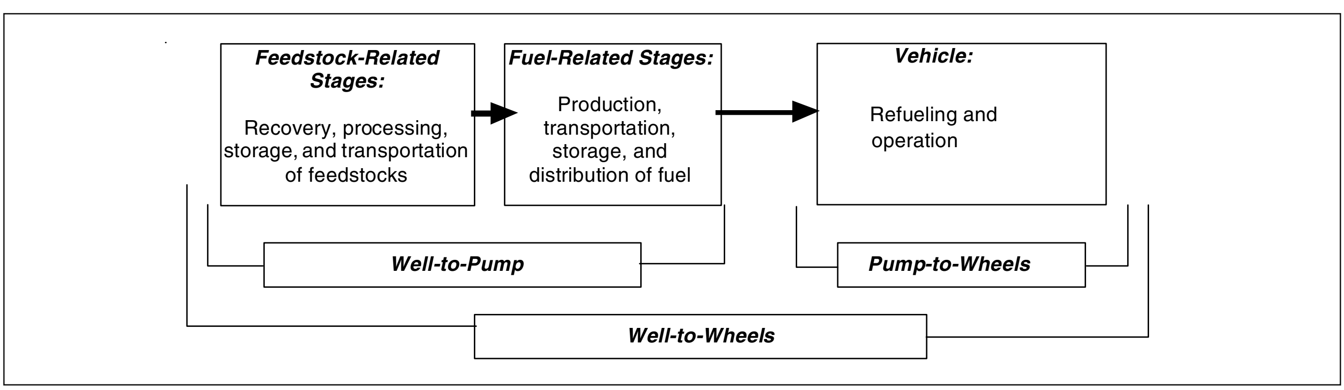

Fig. 1: GREET well-to-wheels modeling boundary for fuel/vehicle systems

1.1 Cell Input

The first button, 'Cell Input', is for the specification of input probability distribution for each uncertain variable. Select one of the parametric assumption cells for which a probability distribution is to be specified and click on 'Cell Input', a pop-up menu is provided. The gallery window contains as many as 11 probability distribution function types. They are: 1) Normal, 2) Lognormal, 3) Uniform, 4) Triangular, 5) Weibull, 6) Beta, 7) Gamma, 8) Extreme Value, 9) Exponential, 10) Pareto, and 11) Logistic. For details on each probability distribution function type please refer to Subramanyan and Diwekar (2005b). The user can select a type of distribution and click 'OK.' An input parameter specification window for the particular distribution opens up. Once a cell has been assigned an input distribution, it turns green.

Fig.3 gives an example input parameter specification window for the GREET default distribution function of a variable, natural gas (NG) recovery efficiency. There are four options in the input specification window:

a) **Input Specification Options Frame**: This option, at the right hand side, consists of radio buttons which can be used to select the type of inputs. As seen from the figure, normal distribution requires two input parameters, which can be selected from one of the following five input specification choices: i) Mean value and standard deviation; ii) 1st and 99th percentile; iii) 5th and 95th percentile; iv) 10th and 90th percentile; and v) 20th and 80th percentile. A 'percentile' can be defined as a score location below which a specified percentage of the population falls. When the inputs are in terms of percentile, the code automatically estimates the values of the mean and standard deviation. When inputs are defined in terms of percentiles, care should be taken to provide feasible percentile values.

b) **Input Parameters Boxes:** These boxes are above the control buttons. Once the type of input parameter is selected, the selected parameter automatically appears beside the input specification boxes. For example, in Fig. 3, the 20th and 80th input specification option has been selected and so they appear as labels of the input text boxes. Here, 0.96, the P20 value, means that there is a probability of 20% that the actual NG recovery efficiency value would be equal to or below 96%.

c) **Minimum and Maximum Cut-off Specification Boxes:** The default minimum and maximum cut-off values, in case of the normal distribution are '–Infinity' and '+Infinity,' respectively. These values are used in case you want to sample from the whole distribution. If you want to truncate the distribution so that samples cannot be less or greater than a particular value, you can truncate the distribution by specifying the particular values in these boxes. For example, the energy efficiency cannot be greater than 1; therefore, the maximum value of the distribution has to be specified as 1 and the plot is truncated at this value (see Fig. 3).

d) **Probability Distribution Function Plot:** Once the input parameters for the probability distribution has been specified, you can visualize the shape of the plot by clicking on the button captioned 'Enter.' The plot is automatically redrawn according to the current input parameters. This is useful if you want to see the variation in the plot for various input parameters. The plot window also has a mean value line that specifies the mean value of the probability distribution function.

The new version, GREET1.7, contains more than 700 default distribution functions for key parameters, such as energy efficiencies, GHG emission factors, as well as criteria pollutant emission factors, for each WTW stage. To accomplish this, the data from each source type were read into Crystal Ball™, a statistical software which, based on the

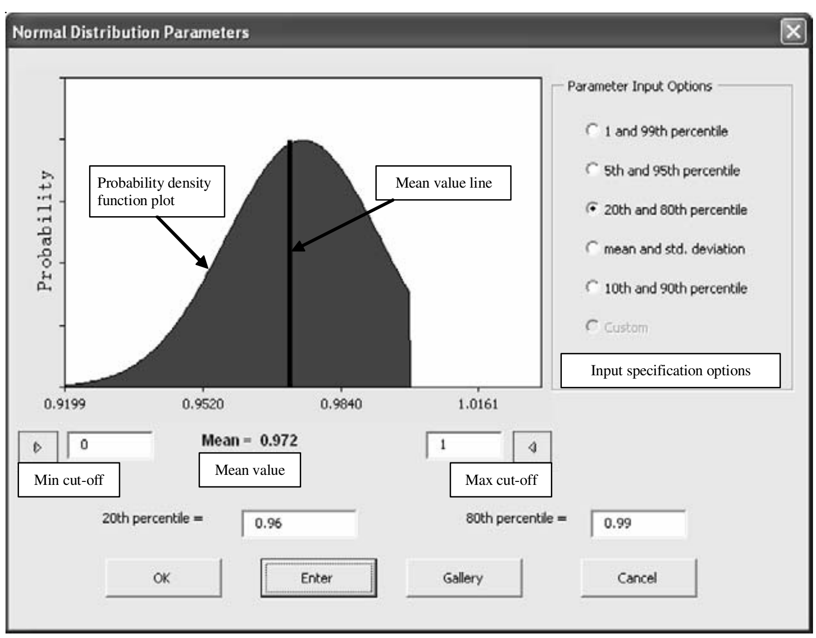

Fig. 3: An example of input specification window for the GREET input variable 'NG recovery efficiency'

number of data points and scatter of the data, attempts to fit a distribution about the data for that source type. In Crystal Ball ™, a mathematical fit is performed to determine the set of parameters for each set of standard distribution functions that best describes the characteristics of the data. The quality or closeness of each fit is determined using a Chi-squared test. All distributions were also visually examined for reasonableness. Ideally, this method should be employed for each cell with distribution function. However, limited data availability often prevented us from taking the statistical approach. In these cases, judgments were made to develop subjective distribution functions. For example, we only have limited data for the key parameter, density of conventional crude oil. In this case, we decided to use triangular distribution function for this parameter with minimum, maximum and most likely values. However, when new data come available, we will employ the statistical approach to improve the distribution quality of these cells. The detailed discussion on the methodology and data sources for the default distribution function database in the GREET model could be found in General Motors Corporation et al. (2001), Wang (2002), Brinkman et al. (2005), and Wu et al. (2006).

1.2 Sampling

Once the distribution functions for all the uncertain parameters have been specified, the next step is to specify the sampling technique to be used and the number of samples required. When the user clicks on 'Sampling' in the stochastic simulation command bar, a window appears. The user can select from one of four sampling techniques: 1) Hammersley Sequence sampling (HSS) (or leaped HSS for dimension >15); 2) Monte Carlo sampling (MCS); 3) Latin Hypercube sampling (LHS); 4) Latin Hypercube Hammersley sampling (LHHS) (or leaped LHHS for dimension >15).

1.2.1 Monte Carlo sampling

One of the most widely used techniques for sampling from a probability distribution is the Monte Carlo sampling technique, which is based on a pseudo-random generator used to approximate a uniform distribution (i.e., having equal probability in the range from 0 to 1). The specific values for each input variable are selected by inverse transformation over the cumulative probability distribution. A Monte Carlo sampling technique also has the important property that the successive points in the sample are independent.

1.2.2 Latin Hypercube sampling

The main advantage of the Monte Carlo method lies in the fact that the results from any Monte Carlo simulation can be treated using classical statistical methods; thus results can be presented in the form of histograms, and methods of statistical estimation and inference are applicable. Nevertheless, in most applications, the actual relationship between successive points in a sample has no physical significance; hence the randomness/independence for approximating a uniform distribution is not critical (Knuth 1973). Moreover, the error of approximating a distribution by a finite sample depends on the equidistribution properties of the sample used for U(0,1) rather than its randomness. Once it is apparent that the uniformity properties are central to the design of sampling techniques, constrained or stratified sampling techniques become appealing (Morgan and Henrion 1990). Latin Hypercube sampling is one form of stratified sampling that can yield more precise estimates of the distribution function. In Latin Hypercube sampling, the range of each uncertain parameter Xi is sub-divided into non-overlapping intervals of equal probability. One value from each interval is selected at random with respect to the probability distribution in the interval. The 'n' values thus obtained for X1 are paired in a random manner (i.e., equally likely combinations) with 'n' values of X2. These 'n' values are then combined with n values of X3 to form n-triplets, and so on, until n k-tuplets are formed. In median Latin Hypercube (MLHS) this value is chosen as the mid-point of the interval. MLHS is similar to the descriptive sampling described by Saliby (1990). The main drawback of this stratification scheme is that, it is uniform in one dimension and does not provide uniformity properties in k-dimensions.

1.2.3 Hammersley Sequence sampling

In late the 1990s, an efficient sampling technique, HSS, based on Hammersley points was developed (Kalagnanam and Diwekar 1997). It uses an optimal design scheme for placing the n points on a k-dimensional hypercube. This scheme ensures that the sample set is more representative of the population, showing uniformity properties in multi-dimensions, unlike Monte Carlo, Latin Hypercube, and its variant, the Median Latin Hypercube sampling techniques. **Fig. 4** graphs the samples generated by different techniques on a unit square. This provides a qualitative picture of the uniformity properties of the different techniques. It is clear from

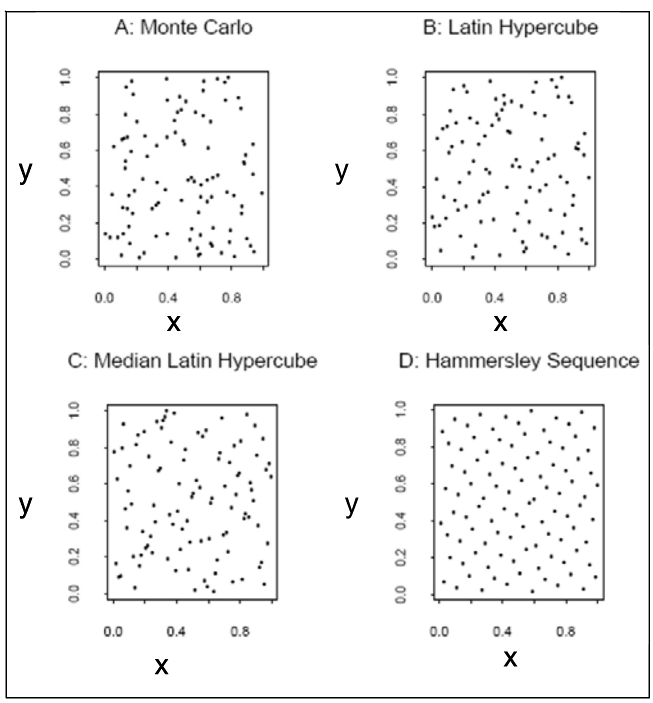

Fig. 4: Sampling points (100) on a unit square using various sampling techniques

Fig. 4 that the Hammersley points have better uniformity properties compared to other techniques. The main reason for this is that the Hammersley points are an optimal design for placing n points on a k-dimensional hypercube. In contrast, other stratified techniques such as the Latin Hypercube are designed for uniformity along a single dimension and then randomly paired for placement on a k-dimensional cube. Therefore, the likelihood of such schemes providing good uniformity properties on high dimensional cubes is extremely small. One of the main advantages of Monte Carlo methods is that the number of samples required to obtain a given accuracy of estimates does not scale exponentially with the number of uncertain variables. HSS preserves this property of Monte Carlo. For correlated samples, the approach uses rank correlations to preserve the stratified design along each dimension. Although this approach preserves the uniformity properties of the stratified schemes, the optimal location of the Hammersley points are perturbed by imposing the correlation structure. Although the original HSS technique designs start at the same initial point, it can be randomized by choosing the first prime number randomly. The HSS technique is much faster than LHS and Monte Carlo techniques and hence is a preferred technique for uncertainty analysis as well as optimization under uncertainty.

1.2.4 Latin Hypercube Hammersley sampling

In this sampling technique, we have used the k-dimensional uniformity of HSS and one dimensional uniformity of LHS to obtain a new sampling technique called Latin Hybercube Hammersley Sampling (Wang et al. 2004). In the process of generating samples with LHHS, the sample values of each input variable are first generated using LHS. The next step is to pair them and combine the input vectors. The conventional method is to pair all of them randomly. However, the sample correlation matrix of input variables generated by either LHS or MCS with random pairing processes is not exactly equal to I and it also shows bad uniformity. Hence restricted pairing procedure is used in all cases. Even when the input variables are independent, the restricted pairing procedure is still employed for the desired correlation matrix I to make sure there is no actual dependence among the input variables. Kalagnanam and Diwekar (1997), Diwekar and Kalagnanam (1996, 1997), and Diwekar (2003b) already showed that Hammersley sequence points have better multidimensional uniformity. In order to characterize the new sampling technique this property, the HSS matrix $\vec{H}(N \times k)$ corresponding to van der Waerden scores matrix in Iman and Conover's approach in LHS, is used in pairing procedures.

To avoid the problem associated with $\vec{H}(N \times k)$ not having a correlation matrix equal to I, the sample correlation matrix $\vec{R}(k \times k)$ associated with $\vec{H}(N \times k)$ is used to find a matrix $\vec{S}$ so that

$$\vec{S}\vec{R}\vec{S}^T = \vec{C} \tag{1}$$

where $\vec{C}$ is the desired sample correlation matrix. The same as above, the Cholesky factorization is used to find a lower triangular matrix $\vec{Q}$ such that

$$\vec{Q}\vec{Q}^T = \vec{R} \tag{2}$$

Therefore, the solution of $\vec{S}$ can be found, which is given by

$$\vec{S} = \vec{P}\vec{Q}^{-1} \tag{3}$$

And correspondingly the transformation factor for the rank matrix is changed to $\vec{S}$ and the rank matrix becomes

$$\vec{H}^* = \vec{H}\vec{S}^T \tag{4}$$

The correlation matrix of $\vec{H}^*$ is exactly equal to the desired correlation matrix $\vec{C}$. The sample can therefore be paired according to the new rank matrix $\vec{H}^*$ rather than $\vec{H}$. In this pairing process, when a correlation structure is not specified, variance of inflation factor (VIF), defined as the largest element on the diagonal of the inverse of the correlation matrix, is computed to detect the large pairing correlations. As the VIF gets much larger than 1, there may be some undesirably large pairing correlations. For VIF large than 10, there can be serious collinearity (Marquardt 1970, Marquardt and Snee 1975). It has been found that the performance of LHHS is most of the time better than HSS. However, unlike MCS or HSS, the performance measure for LHHS is not independent of number of variables or type of functionality used to compute the output distributions.

1.2.5 Leaped HSS and LHHS

It has been recently found that the uniformity property of HSS for higher dimensions (more than 30 uncertain variables) gets distorted. HSS and LHHS are generated based on prime numbers as bases. In order to break this distortion, we introduced leaps in prime numbers for higher dimensions. This leaped HSS and LHHS showed better uniformity than the basic HSS and LHHS. For simplicity, we have leaped HSS and LHHS as a part of the HSS and LHHS techniques in the GREET stochastic modeling capability. When the number of uncertain variables exceeds 15, the switch occurs automatically.

1.3 Forecast cells

The next step is to select those variables whose values will be forecasted. GREET1.7 includes more than 90 fuel production pathways and more than 70 vehicle/fuel systems (Wang et al. 2005). Therefore, the user can have as many as approximately 3,000 forecast variables for stochastic simulation. A special algorithm has been created to enable you to easily select the forecast variables for the pathways of interest through four simple steps:

a) Select the vehicle technologies.
b) Specify the transportation fuels.
c) Specify the well-to-wheels (WTW) simulations and/or well-to-pump (WTP) simulations.
d) Select the energy and emission forecast groups.

The naming convention for the forecast variables is 'Vehicle Technology – Transportation Fuel – WTW and/or WTP – Energy and Emission Forecast'. For example, 'CIDI-DME-WTW-N2O' can be interpreted as the well-to-wheels total N_2O emissions for the compression-ignition direct-injection (CIDI) vehicle fueled with dimethyl ether (DME). The forecast variables listed in the list box titled 'Selected Forecasts' are those which would be predicted at the end of the stochastic simulation.

1.4 Delete distributions

For any parametric assumption cell with a probability distribution, if the user decides to just assign a point value to that cell, the probability distribution can be deleted by selecting the cell and clicking on the 'Deleted Distribution' button. The input distribution is automatically deleted and the cell color turns from green to white.

1.5 Run simulation

After all the required inputs and forecast selections have been finalized, the 'Run Simulation' button is enabled to click to begin execution of the stochastic simulation. After the simulation run is completed, the forecast results are exported to another Excel file and statistical values like the mean, standard deviation, and 0^{th} to 100^{th} percentile are calculated automatically for each forecast variable.

2 Results and Discussion

The fuel-cell vehicle (FCV) fueled with liquid hydrogen (LH2) was chosen for a case study to evaluate the performance of the four sampling techniques. The stochastic simulation was run with each sampling technique and the output results were analyzed and compared. **Table 1** presents the central processing unit (CPU) times taken for 16 independent runs varying the number of samples and the sampling technique.

Table 1: CPU time (in hrs.min.sec) for each stochastic simulation on a 2.39GHz processor

Sampling techniques	Number of samples			
	1000	2000	3000	4000
HSS	0.22.38	0.44.57	1.07.13	1.31.02
MCS	0.22.00	0.45.17	1.08.23	1.29.41
LHS	0.21.31	0.41.56	1.03.21	1.25.09
LHHS	0.21.22	0.41.47	1.02.50	1.24.13

Table 2: Mean and standard deviation (S.D.) values of forecast variable 'WTW total energy use of LH2 FCV' (1000 samples)

Sampling techniques	Total energy use: Btu/mi	
	Mean	S.D.
HSS	5853.98	1029.25
MCS	5837.51	1047.53
LHHS	5857.13	1029.63
LHS	5858.70	1050.17

Figs. 5 (a) through (d) show the output distribution curves for a particular forecast variable which is the WTW total energy use of LH2 FCV from four stochastic runs each with 1000 samples. The four charts correspond to MCS, HSS, LHHS and LHS, respectively. The comparison of the output distribution curves substantiates the uniformity comparisons in Fig. 4. The CPU times for all these four runs are almost equal as seen from Table 1 but the HSS curve is the smooth-

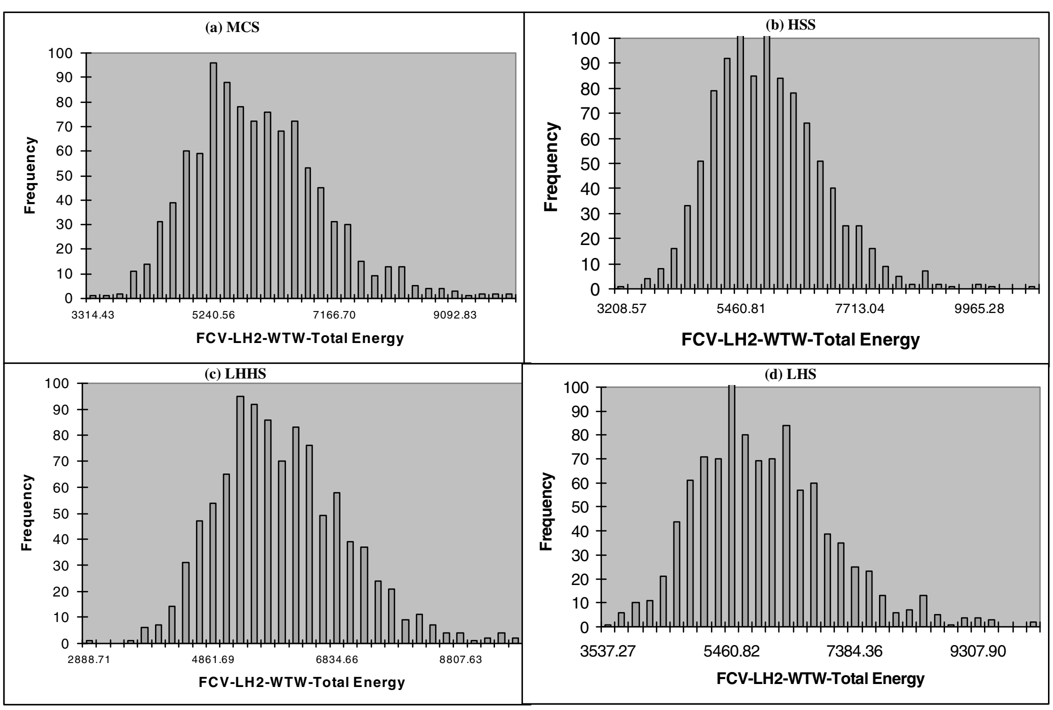

Fig. 5: The output distribution curves for the forecast variable 'WTW total energy use of LH2 FCV' with 1,000 samples using: a) MCS; b) HSS; c) LHHS; and d) LHS

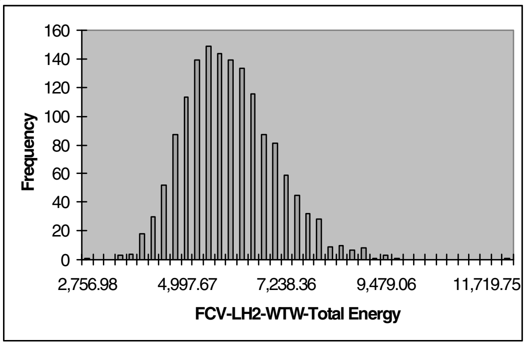

Fig. 6: The output distribution curves for the forecast variable 'WTW total energy use of LH2 FCV' using LHHS with 1,500 samples

est corresponding to the highest level of uniformity, followed by LHHS and while MCS and LHS are almost the same with respect to smoothness factor. **Table 2** lists the mean values and the standard deviation of all four sampling techniques for this forecast. The interesting fact here is that the means from all sampling techniques are within 0.3% of each other and the standard deviation of the HSS and LHHS techniques are only 2% lower than the other two. Therefore, even though the mean values and variances are all close to each other, the Figs. 5 (a)–(d) show that 1,000 samples are optimal for HSS to obtain a relatively smooth curve but not for MCS, LHHS and LHS.

Fig. 6 illustrates the output distribution curve as the number of samples for LHHS was increased to 1,500 with all other parameters remaining the same. It can be clearly seen that the curve is considerably smoother than the 1,000 sample case (see Fig. 5 (c)) though at a cost of CPU time which has increased to about 33 min. The mean value and standard deviation with 1,500 samples were computed as 5856.38 and 1028.42, respectively, with the differences both less than 0.1% compared to the results with 1,000 samples.

Fig. 7 (**a**) illustrates the output distribution curve as the number of samples for LHS was increased to 2,000 with all other parameters remaining the same. As expected, the smoothness of the curve has improved with the increase in number of samples though still not as smooth as that of HSS or LHHS. This means that LHS needs more samples to reach the smoothness level of HSS and LHHS, for example, 3,000 samples (see **Fig. 7** (**b**)) or more. However, the CPU time has increased significantly to 42 min for 2,000 samples and 63 min for 3,000 samples (see Table 1). Therefore, the user needs to assess the trade off to balance the number of samples and simulation time for LHS. Again, the computed mean values and standard deviation values for 2,000 and 3,000 samples show negligible effects on the mean and variance no matter the samples are doubled or tripled.

As a test, we did LHS and MCS simulations with 4,000 samples (the distribution curves are not shown in the paper). Our finding is the marginal improvement in smoothness of the output distribution curve from 3,000 samples to 4,000 samples is not as much as that from 1,000 samples to 2,000 samples and 2,000 samples to 3,000 samples. However, the time of CPU is reaching as long as one and a half hour if the user selects 4,000 samples. We noticed the increasing number of samples did make the simulated mean value marginally closer to the actual mean value; however, this improvement is so tiny and does not deserve the extension of simulation time.

With this stochastic simulation case study, we applied HSS as the default technique with 1,000 as the default sampling number in GREET according to the performance comparison among these four techniques. If the user picks LHHS technique for stochastic simulation, the optimal number of samples is 1,500. In case of LHS technique, there is a clear difference in the smoothness between the 2,000 and 3,000 sample cases. But this difference costs 22 min longer of CPU time for the 3,000 sample case. This trade-off needs to be carefully assessed by the user based on his/her preferences. Similar to LHS, an acceptable level of smoothness was reached with ~3,000 samples for the MCS technique. Therefore, HSS has achieved more than 200% reduction in running time without compromising on the accuracy and quality of the prediction curves while LHHS has achieved a 150% reduction. This is due to the fact that for both techniques, a lesser number of samples are required for representing the sample space with substantial uniformity.

3 Conclusions

A new stochastic simulation tool has been developed to be built in the Argonne's GREET model to enhance the capability for addressing the uncertainties incorporated in a wide variety of input parameters. This tool has been designed as

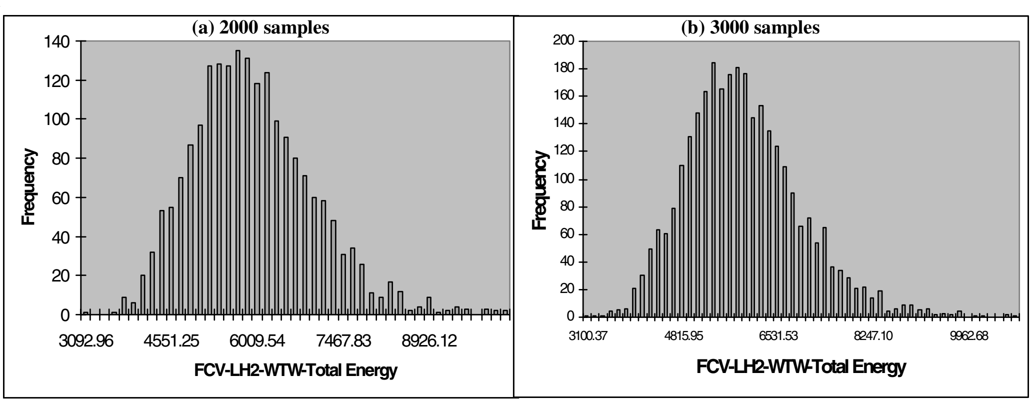

Fig. 7: The output distribution curves for the forecast variable 'WTW total energy use of LH2 FCV' using LHS with: a) 2,000 samples; and b) 3,000 samples

a Microsoft® Excel add-in file with Visual Basic macros which can be loaded whenever the user needs to perform a stochastic simulation within the model. This new tool automates the process of setting up a stochastic simulation to a great extent and guides the user in each step of the process through the user-friendly graphical user interface (GUI) windows.

In general, totally five steps should be followed to perform a complete stochastic simulation through the new tool:

a) **Cell Input:** Specify probability distribution functions to the input variables;
b) **Sampling:** Indicate the number of samples required and the sampling technique to be used;
c) **Forecast Cells:** Define the forecast variables;
d) **Delete Distribution:** To delete a distribution defined previously;
e) **Run Simulation:** Propagate the uncertainties and statistically analyze the outputs.

The new version, GREET1.7, includes more than 90 fuel production pathways and more than 70 vehicle/fuel systems. As a result, the model contains more than 700 default distribution functions for various key parameters, such as energy efficiencies, GHG emissions factors, as well as criteria pollutant emission factors, for each WTW stage; and as many as ~3,000 forecast variables for stochastic simulation. During the stochastic simulation, the selection of sampling technique is a key factor. The new tool incorporates four sampling techniques for GREET stochastic simulations: Monto Carlo sampling, Latin Hypercube sampling, Hammersley Sequence sampling, and Latin Hypercube Hammersley sampling.

In this paper, we did 16 independent stochastic simulations in the GREET model to evaluate the performance of the four sampling techniques. The main findings are as follows:

a) Despite the selection of sampling technique and the number of samples (between 1,000 and 4,000), the simulated mean values are close to the actual mean value enough (within ±1%). The standard deviation values from each other are close enough as well (within ±5%). It shows the trend that the increasing number of samples makes the simulated mean value marginally closer to the actual mean value; however, the improvement effect is negligible.
b) With the same number of samples, the output distribution curve simulated by HSS is the smoothest corresponding to the highest level of uniformity, followed by LHHS and while MCS and LHS are almost the same with respect to smoothness factor.
c) To achieve the same level of smoothness as HSS with 1,000 samples, LHHS needs to be simulated with ~1,500 samples and LHS and MCS with ~3,000 samples. Because the simulation time is almost the same for running each sample by these four techniques, HSS can achieve more than 200% reduction in running time without compromising on the accuracy and quality of the prediction curves while LHHS can achieve a 150% reduction compared to LHS or MCS.
d) The simulation time is strictly positive-correlated with the number of samples; therefore, the trade-off of extending simulation time and improving smoothness of output distribution curve needs to be carefully assessed by the user based on his/her preferences.

According to the performance comparison among these four techniques, we applied HSS as the default technique with 1,000 as the default sampling number in GREET.

Acknowledgements. This work was funded by Argonne National Laboratory. Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC, under contract No. DE-AC02-06CH11357. The authors thank Dr. Amgad Elgowainy of Argonne National Laboratory for his inputs to this work.

References

Brinkman N, Wang M, Weber T, Darlington T (2005): Well-to-wheels analysis of advanced fuel/vehicle systems – A North American study of energy use, greenhouse gas emissions, and criteria pollutant emissions. <http://www.transportation.anl.gov/software/GREET/publications.html>

Diwekar U (2003a): Introduction to applied optimization. Kluwer Academic Publishers, Dordrecht

Diwekar U (2003b): A novel sampling approach to combinatorial optimization under uncertainty. Computational Optimization and Applications 24, 335–371

Diwekar U, Kalagnanam J (1996): Robust design using an efficient sampling technique. Computers and Chemical Engineering 20 (Suppl) S389–S394

Diwekar U, Kalagnanam J (1997): An efficient sampling technique for optimization under uncertainty. AIChE Journal 43, 440–459

General Motors Corporation, Argonne National Laboratory, BP, ExxonMobil, Shell (2001): Well-to-wheels energy use and greenhouse gas emissions of advanced fuel/vehicle systems – North American analysis. <http://www.transportation.anl.gov/software/GREET/publications.html>

Kalagnanam J, Diwekar U (1997): An efficient sampling technique for off- line quality control. Technometrics 39 (3) 308–319

Knuth D (1973): The art of computer programming, Volume 1: Fundamental algorithms. Addison-Wesley Publishing Co.

Marquardt D (1970): Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12, 591–612

Marquardt D, Snee R (1975): Ridge regression in practice. The American Statistician 29, 3–20

Morgan G, Henrion M (1990): Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge, UK

Saliby E (1990): Descriptive sampling: A better approach to Monte Carlo simulations. J Operations Research Society 41 (12) 1133

Subramanyan K, Diwekar U (2005a): Characterization and quantification of uncertainty in solid oxide fuel cell hybrid power plants. J Power Sources 142, 103–116

Subramanyan K, Diwekar U (2005b): User manual for stochastic simulation capability in GREET. <http://www.transportation.anl.gov/software/GREET/index.html>

Wang M (1996): Development and use of the GREET model to estimate fuel-cycle energy use and emissions of various transportation Technologies and fuels. Center for Transportation Research, Argonne National Laboratory, ANL/ESD-31

Wang M (2002): Fuel choices for fuel-cell vehicles: Well-to-wheels energy and emission impacts. J Power Sources 112, 307–321

Wang M, Wu Y, Elgowainy A (2005): Operating manual for GREET: Version 1.7. Center for Transportation Research, Argonne National Laboratory, ANL/ESD/05-3

Wang R, Diwekar U, Gregoire-Padro C (2004): Latin Hypercube Hammersley sampling for risk and uncertainty analysis. Environmental Progress 23 (2) 141

Wu Y, Wang M, Sharer P, Rousseau A (2006): Well-to-wheels results of energy use, greenhouse gas emissions, and criteria air pollutant emissions of selected vehicle/fuel systems. In Journal of Engines, SAE 2006 Transactions, SAE 2006 World Congress, 2006-01-0377, Detroit, MI

Received: February 15th, 2007
Accepted: July 24th, 2007
OnlineFirst: July 25th, 2007